

UNIVERSITAS BINA NUSANTARA

Program Ganda
Teknik Informatika - Matematika
Skripsi Sarjana Program Ganda
Semester Ganjil 2005/2006

SKRIPSI PROGRAM GANDA UNIVERSITAS BINA NUSANTARA

Sintiche Mayang Suwandi
0500586733

Abstrak

Masalah yang dihadapi dalam melakukan proses pencarian dokumen pada *search engine* adalah kurangnya keterkaitan antara dokumen–dokumen hasil pencarian. Hasil yang ditampilkan oleh *search engine* harus dapat mengidentifikasi kelompok dari dokumen yang ingin dicari secara spesifik. Masalah seperti ini belum cukup terpecahkan dalam pencarian sebelumnya, di mana hasil yang didapatkan terlalu panjang, ambigu dan munculnya dokumen-dokumen yang tidak berkaitan. Oleh karena itu penulis merasa perlu untuk mengusulkan alternatif metode lain, yang dapat meminimalisasi usaha *user* dalam mencari dokumen yang diinginkan.

Adapun metode yang digunakan untuk meminimalisasi usaha *user* dalam mencari dokumen yang diinginkan yaitu dengan menggunakan algoritma Lingo. Algoritma Lingo memiliki kelebihan berupa pengembalian hasil pencarian dokumen dalam bentuk pengelompokan dan urutan relevansi yang jelas.

Dengan menggunakan *search engine* yang ada saat ini cukup memakan waktu dalam menemukan dokumen yang tepat sesuai dengan keinginan *user*, karena hasil pencarian berupa daftar panjang dari dokumen tanpa pengelompokan dan tingkat relevansi yang jelas. Sedangkan dari alternatif sistem yang diusulkan penulis berhasil meminimalisasi usaha *user* dalam mencari dokumen yang diinginkan dengan melakukan pengembalian hasil pencarian berupa daftar dokumen dalam bentuk pengelompokan dan urutan relevansi yang jelas. Oleh karena itu perlu adanya pengkajian kembali terhadap sistem dengan cara cara merubah pola pikir para pengelola *search engine* yang ada mengenai kemudahan pencarian dokumen bagi *user* dengan teknologi baru.

Kata Kunci:

algoritma Lingo, *information retrieval*, *stemming*, *stop words removal*, *Singular Value Decomposition*.

PRAKATA

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi ini. Skripsi ini disusun sebagai salah satu syarat dalam menyelesaikan program studi ganda Strata 1 di Universitas Bina Nusantara, Jakarta.

Selama menyelesaikan skripsi ini, penulis mendapat banyak bantuan serta dukungan moral dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terimakasih yang sebesar-besarnya kepada :

1. Bapak Prof. Dr. Gerardus Polla, M.AppSc. selaku rektor Universitas Bina Nusantara.
2. Bapak Wikaria Gazali, S.Si, M.T., selaku Dekan Fakultas MIPA Universitas Bina Nusantara sekaligus sebagai dosen pembimbing untuk waktu, sumbangan pikiran, dan bimbingannya dalam menyelesaikan masalah yang ditemui penulis selama menyusun skripsi ini.
3. Bapak Ngarap Imanuel Manik, Drs., M.Kom., selaku Ketua Jurusan Matematika Universitas Bina Nusantara.
4. Bapak Dr. Ir. Haryono Soeparno, M.Sc selaku dosen pembimbing untuk waktu, sumbangan pikiran, dan bimbingannya dalam menyelesaikan masalah yang ditemui penulis selama menyusun skripsi ini.
5. Orang tua penulis yang telah memberikan dukungan yang sangat besar baik moral maupun finansial untuk menyelesaikan skripsi ini dengan baik.
6. Teman-teman kuliah penulis dan teman-teman Binus Center untuk setiap dukungannya.
7. Pihak-pihak lain yang tidak dapat penulis sebutkan satu persatu yang telah membantu penulis dalam memberikan dukungan secara langsung maupun tidak langsung dalam penulisan skripsi ini

Penulis menyadari bahwa skripsi ini masih belum sempurna. Oleh karena itu, penulis sangat menyambut baik kritik dan saran yang membangun dari semua pihak. Semoga skripsi ini dapat memberikan manfaat yang besar bagi para pembaca.

Jakarta, Januari 2006

Penulis

Sintiche Mayang Suwandi

0500586733

DAFTAR ISI

	Halaman
Halaman Judul Luar	i
Halaman Judul Dalam	ii
Halaman Persetujuan Hardcover	iii
Halaman Pernyataan Dewan Penguji	iv
Abstrak	v
Prakata	vi
Daftar Isi	vii
Daftar Gambar	ix
Daftar Tabel	x
Daftar Lampiran	xi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Ruang Lingkup	2
1.3 Tujuan dan Manfaat	3
1.4 Metodologi	4
1.5 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	6
2.1 Kerangka Teori	6
2.1.1 Algoritma Lingo	6
a <i>Preprocessing</i>	9
1 Metode <i>Stop Words Removal</i>	9
2 Metode <i>Stemming</i>	10
b Metode <i>Frequent Phrase Extraction</i>	14
c Metode <i>Cluster Label Induction</i>	18
d Metode <i>Cluster Content Discovery</i>	24
2.1.2 Matriks Ortogonal	27

2.2	Kerangka Pikir	27
BAB 3	ANALISA DAN PERANCANGAN	31
3.1	Gambaran Umum	31
3.2	Alternatif Pemecahan Masalah	32
3.3	Perancangan Program Aplikasi	34
3.3.1	Perumusan Objek Penelitian	34
3.3.2	Data Yang Digunakan	35
3.3.3	Perancangan Program Aplikasi	37
3.3.4	Diagram Alur Tahapan Algoritma Lingo	41
BAB 4	IMPLEMENTASI DAN EVALUASI	45
4.1	Spesifikasi Sistem	45
4.2	Evaluasi Program Aplikasi	46
4.2.1	Input Data	46
4.2.2	Tampilan Layar	47
BAB 5	KESIMPULAN DAN SARAN	69
5.1	Kesimpulan	69
5.2	Saran	70
	DAFTAR ACUAN	71
	DAFTAR PUSTAKA	71
	RIWAYAT HIDUP	72
	LAMPIRAN	L1

DAFTAR GAMBAR

	Halaman
Gambar 2.1 http://citeseer.ist.psu.edu/osinski04conceptual.html	6
Gambar 3.1 Rancangan Tampilan Layar	41
Gambar 3.2 Diagram Alur <i>Preprocessing</i>	42
Gambar 3.3 Diagram Alur <i>Frequent Phrase Extraction</i>	43
Gambar 3.4 Diagram Alur <i>Cluster Label Induction</i>	43
Gambar 3.5 Diagram Alur <i>Cluster Content Discovery</i>	44
Gambar 3.6 Diagram Alur Tahapan Algoritma Lingo	44
Gambar 4.1 Layar Awal	47
Gambar 4.2 Tampilan Layar Setelah Penekanan Tombol “Baca Seluruh File”	48
Gambar 4.3 Tampilan Layar Setelah Penekanan Tombol “Baca Daftar Stop Words”	49
Gambar 4.4 Tampilan Layar Setelah Penekanan Tombol “Buang Kata Stop Words”	50
Gambar 4.5 Tampilan Layar Setelah Penekanan Tombol “Stemming”	51
Gambar 4.6 Tampilan Layar Setelah Penekanan Tombol “Buat Vektor Dokumen dan Normalisasi”	52
Gambar 4.7 Tampilan Layar Setelah Penekanan Tombol “SVD”	53
Gambar 4.8 Tampilan Layar Setelah Penekanan Tombol “Hitung Banyaknya Kelompok”	54
Gambar 4.9 Tampilan Layar Setelah Penekanan Tombol “Buat Vektor Frase dan Normalisasi”	55
Gambar 4.10 Tampilan Layar Setelah Penekanan Tombol “Hitung Deskripsi Kelompok”	56
Gambar 4.11 Tampilan Layar Setelah Penekanan Tombol “Penandaan File Ke Kelompok”	57

DAFTAR TABEL

		Halaman
Tabel 2.1	Daftar kata yang termasuk <i>Stop Words Removal</i>	10
Tabel 4.1	Hasil Program Aplikasi pada proses <i>Preprocessing</i>	59

DAFTAR LAMPIRAN

		Halaman
Lampiran 1	Contoh Soal Dalam Memperoleh Matriks Ortogonal	L1
Lampiran 2	<i>Pseudocode</i> Algoritma Lingo	L6
Lampiran 3	Contoh Soal Dalam Pengerjaan Langkah Algoritma Lingo	L8
Lampiran 4	Lampiran <i>Coding</i>	L12